YO'I'O
You only Invest Once

Recommend stocks by analyzing Reddit's content and topic modelling (BERT)

Unstructured Data Analytics Project | Fall 2022
McCombs School of Business, UT Austin

Kavya Angara
Prathmesh Savale
Pratik Gawli
Rajshree Mishra
Aniket Patil

Why content analysis matters while recommending stocks ?

- Financial analyst working at a YO'I'O investment company
- Importance of understanding people's sentiments about the market along with company performance necessary
- Recently becoming an important factor in deciding upon which stocks should the investment company track
- Ultimately decide whether one should buy a particular stock/stocks from a particular industry to maximize the profits.
- **Wallstreetbets** on Reddit is one of the many public forums where people discuss such recent market trends and express their sentiments about them

Road Map

Scraping data from Reddit → Topic Modelling → Sentiment Analysis → Recommend Stocks

Data Source

- **For Topic Modelling**
  - Scraped ~500k comments & posts from the subreddit r/Wallstreetbets
  - Date range: Sept 1, 2022 to Sept 30, 2022
- **For Sentiment Analysis**
  - Scrape data to get top stock tickers from Sept 1, 2022 to Sept 30, 2022
  - Scrape data for only those top tickers to perform sentiment analysis
- **Stock Simulator - Retrospect**
  - Scraped Yahoo finance for actual monthly closing prices

Pre-processing Steps

- **For Topic Modelling**

  - Length of comment more than 30 words

  - Remove outlier topics (taken care by BERT)

- **For Sentiment Analysis**

  - Assign sentiment score on lexicons we got from topic modelling

  - +2 for positive sentiment & -10 for negative sentiment

- **Stock Simulator - Retrospect**

  - Perform sentiment analysis on top 10 tickers in windows of 90, 60 and 30 days and

  - Get actual data from yahoo finance for the same top 10 tickers to compare our recommendations
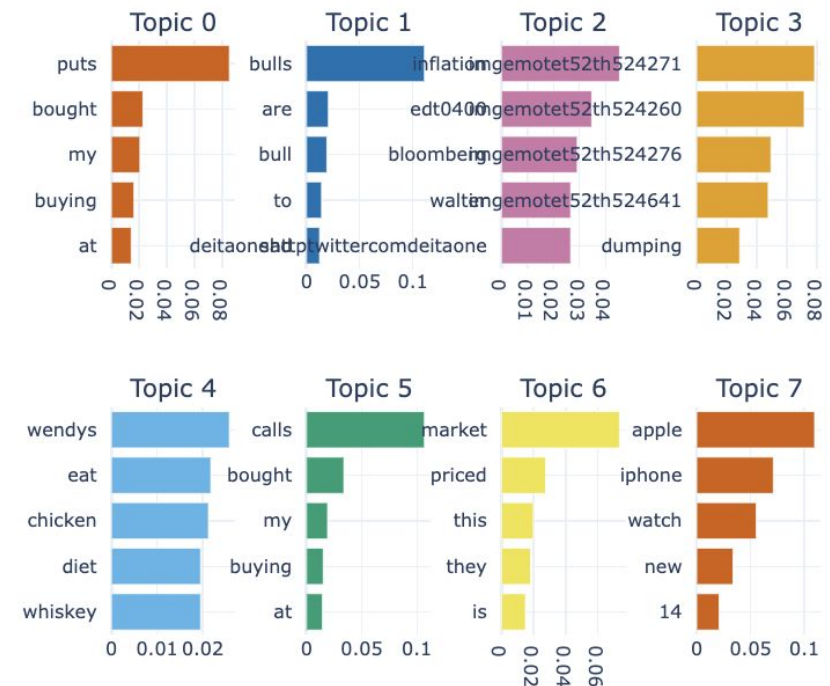
Topic Modelling | Analysis & Results

We have used **BERTtopic** for topic modelling

to understand the context of the market.

The top 5 topics based on the forum data are:

- Buying puts
- SPY down
- Bulls are sad
- Inflation
- Money Trading

This is as expected because from 1st September to 30th September the market (S&P 500 index) has gone down from 3966.85 to 3585.62.
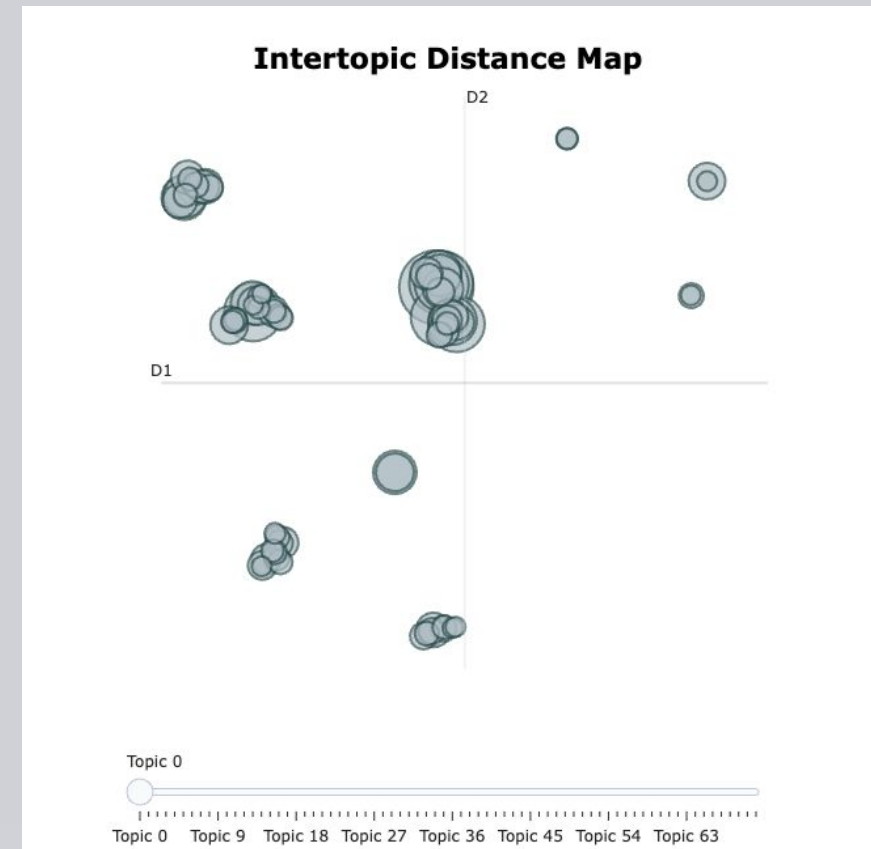

Topic Word Scores

Topic Modelling | Analysis & Results

Could topics like Wendy's to earn money, queen's demise, divorces & girlfriend be related to the main topic of discussion about stocks?
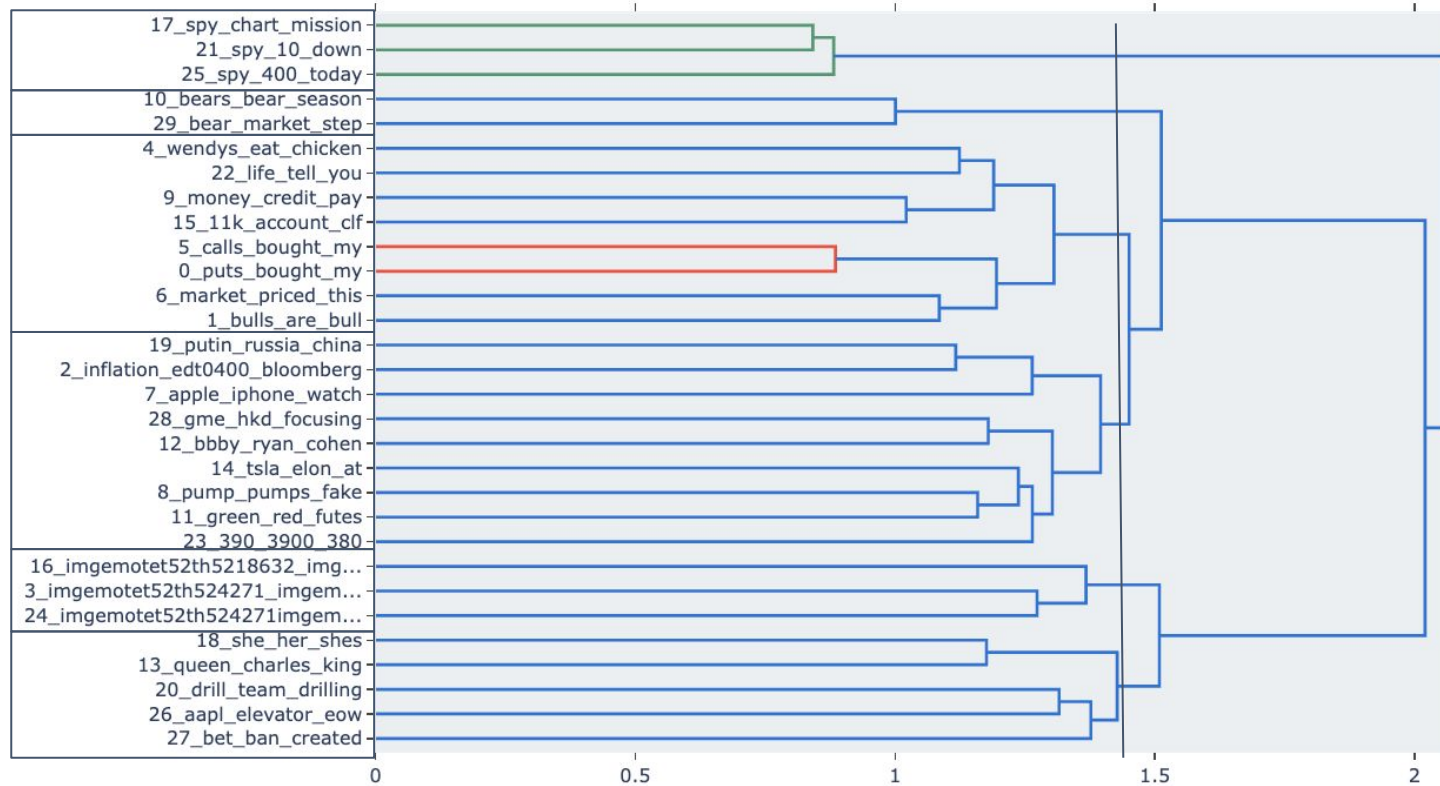
Central cluster (largest cluster) includes most of the major topics relating to discussions about the market going down

The second largest cluster (to the left of the central cluster) includes topics about union strikes, which could be triggered by inflation, a main cause of the market falling

Topic Modelling | Analysis & Results

Hierarchical Clustering

Sentiment Analysis | Analysis & Results

- Topic modelling tells us that there is a strong negative sentiment in the market, people talk about buying puts, inflation, working at Wendys, filing for divorces and leaving their girlfriends

- We capture this sentiment by adding lexicons for the WSB specific slang terms in vader, weighting negative words more than positive

- Sentiment analysis is run for the most frequently discussed stocks

Results/Scores for the 5 most discussed tickers:

	ticker	tf	compound_sentiment
0	GME	382	0.029451
2	AI	127	-0.022523
3	BBBY	101	-0.026281
5	AMC	90	-0.049026
6	CS	85	-0.111998

modified rules for VADER:

Word	Assigned Score
tendies	2
bull	2
calls	2
moon	2
btfd	2
bear	-10
short	-10
sell	-10
puts	-10
going t*** up	-10

Recommendations and Stock Simulator – Retrospective

Compound Sentiment Score from VADER

ticker	compound_sentiment_2months_back	compound_sentiment_1month_back
GME	-0.035763	-0.034960
AI	0.144990	0.094232
BBBY	-0.000798	0.007149
AMC	0.056591	0.004049
CS	0.096831	0.149132
TSLA	-0.052972	-0.082407
IT	-0.036135	-0.041592
NVDA	0.043307	-0.078903
GM	-0.107337	0.184702
AMD	-0.009890	-0.032588

Monthly closing prices (Yahoo Finance)

NaT	2022-07-01	2022-08-01
GME	34.009998	28.639999
AI	18.410000	18.000000
BBBY	5.030000	9.530000
AMC	14.560000	9.120000
CS	5.800000	5.160000
TSLA	297.149994	275.609985
IT	265.480011	285.320007
NVDA	181.630005	150.940002
GM	36.259998	38.209999
AMD	94.470001	84.870003

r/wallstreetbets · Posted by u[...] just now

GME YOLO update — Apr 16 2021 — final update

YOLO

Symbol ▾	Actions	Last Price $	Change $	Change %	Qty #	Price Paid $	Day's Gain $	Total Gain $	Total Gain %	Value $
> GME ⓘ	🔔	154.69	-1.75	-1.12%	200,000	55.1774	-273,366.85	19,902,513.83	180.35%	30,938,000.00
> Cash Total Transfer money										$3,535,248.01
Total						$11,035,486.17	-$273,366.85	$19,902,513.83	180.35%	$34,473,248.01

Thank You -
Diamond Hands Forever